**AP STAT- Ch. 3 -- 5 Quiz Review**

1. A survey of automobiles parked in the student and staff lots at a large university classified the brands by country of origin, as seen in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Driver | |  |
|  |  | **Student** | **Staff** | TOTAL |
|  | **American** | 107 | 105 | 212 |
| Origin | **European** | 33 | 12 | 45 |
|  | **Asian** | 55 | 47 | 102 |
|  | TOTAL | 195 | 164 | 359 |

1. What is the marginal distribution of Origin? Make a bar graph.

American = 59.05%

European = 12.53%

Asian = 28.41%

1. What is the marginal distribution of Driver? Do not make a bar graph.

Student = 54.317%

Staff = 45.682%

1. What percent of Students drove Asian cars?

P(Asian|Student) = 107/195 = 54.872%

1. What percent of Asian cars are driven by staff?

P(Staff|Asian) = 47/102 = 46.078%

1. What percent of Staff drove Asian cars?

P(Asian|Staff) = 47/164 = 28.659%

1. What percent of those surveyed were Students?

P(Students) = 54.317%

1. What percent of those surveyed drove American cars or were students?

P(American U Students) = (55+33+107+105)/359 = 83.565%

1. What percent of those surveyed drive European cars and were staff?

P(European n Staff) = 12/359 = 3.343%

1. What is the conditional distribution of Origin?

**American European Asian**

Student 50.472% 73.333% 53.922%

Staff 49.528% 26.667% 46.078%

1. What is the conditional distribution of Driver?

**Student Staff**

American 54.872% 64.024%

European 16.923% 7.317%

Asian 28.205% 28.659%

1. Create a segmented bar chart for the conditional distribution of Driver.
2. Is there an association between Origin and Driver? Provide statistical evidence to support your claim.

There DOES appear to be an association (the variables appear to be DEPENDENT). This is shown in the stacked bar graph above. There appear to be DIFFERENT percentages of car origins for the different types of drivers. Staff seems to drive more American cars and less European cars than Students do. However it seems that both Students and Staff drive the same percentage of Asian cars.

1. Create a dotplot of the number of goals scored by each team in the first round of the California high school soccer playoffs. Then briefly describe the distribution.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | 0 | 1 | 0 | 7 | 2 | 1 | 0 | 4 | 0 | 3 | 0 | 2 | 0 |
| 3 | 1 | 5 | 0 | 3 | 0 | 1 | 0 | 1 | 0 | 2 | 0 | 3 | 1 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x |  |  |  |  |  |  |  |
| x |  |  |  |  |  |  |  |
| x |  |  |  |  |  |  |  |
| x |  |  |  |  |  |  |  |
| x |  |  |  |  |  |  |  |
| x | x |  |  |  |  |  |  |
| x | x |  |  |  |  |  |  |
| x | x |  | x |  |  |  |  |
| x | x | x | x |  |  |  |  |
| x | x | x | x |  | x |  |  |
| x | x | x | x | x | x |  | x |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

**Number of Goals**

SHAPE: unimodal, right skewed

CENTER: Median of 1 goal

SPREAD: (0, 7)

1. Create back-to-back stemplots of the following male and female heights. Compare & describe both distributions

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **MALE** | |  |  |  |  |  | **FEMALES** | | |  |  |  |
| 72 | 75 | 66 | 76 | 70 | 71 |  | 72 | 69 | 70 | 64 | 70 | 66 |
| 73 | 74 | 65 | 73 | 73 | 66 |  | 70 | 60 | 71 | 65 | 61 | 67 |
| 73 | 68 | 65 | 63 | 72 | 68 |  | 66 | 59 | 70 | 66 | 69 | 68 |
| 70 | 68 | 70 | 64 | 72 | 69 |  | 61 | 61 | 60 | 66 | 68 | 68 |
| 71 | 67 | 71 | 60 | 71 | 72 |  | 60 | 62 | 61 | 66 | 67 | 65 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | **MEN** | |  | **WOMEN** | | | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **5** | 9 |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  | 5 | 5 | 4 | 3 | 0 | **6** | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 4 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | 9 | 8 | 8 | 8 | 7 | 6 | 6 | **6** | 5 | 5 | 6 | 6 | 6 | 6 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 |
| 4 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | **7** | 0 | 0 | 0 | 0 | 1 | 2 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | 6 | 5 | **7** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**SHAPES**: both mens and womens distributions are unimodal. Mens distribution is left skewed while women’s distribution is roughly symmetric.

**CENTERS:** Men’s center is the median of 70.5 which is higher than the women’s mean is 65.6.

**SPREAD:** The men’s spread is (60, 76) which is similar in spread to the women’s spread of (59, 72).

1. Find the 5# summaries and create parallel boxplots for the heights of males and females in question #3

**MEN: WOMEN:**

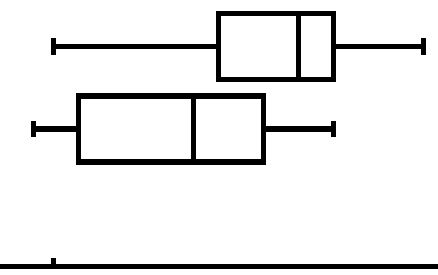
Min: 60 Min: 59

Q1: 67 Q1: 61

Med: 70.5 Med: 66

Q3: 72 Q3: 69

Max: 76 Max: 72



MEN

WOMEN

60 66 69 72 76

1. Salaries of 2008 New York Yankees (in millions of dollars):

Rodriguez 28 Giambi 23.428

Jeter 21.6 Abreu 16

Petite 16 Rivera 15

Posada 13.1 Damon 13

Matsui 13 Mussina 11.071

Pavano 11 Farnsworth 5.917

Wang 4 Hawkins 3.75

Cano 3 Molina 1.875

Ensberg 1.75 Brackman 1.185

Betemit 1.165 Bruney 0.725

Traber 0.500 Cabrera 0.461

Hughes 0.406 Duncan 0.398

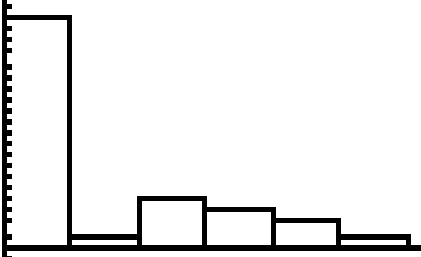
Henn 0.397 Kennedy 0.394

Karstens 0.393 Albaladejo 0.393

Ohlendorf 0.391 Chamberlain 0.390

Sanchez 0.390

1. Create a frequency histogram of the data above. Describe the distribution.



**#**

20

4

1

0 5 10 15 25 30

SALARIES

Shape: Right skewed, unimodal

Center: Median of 1.75 million dollars

Spread: range of (0.39, 28) and an IQR of 12.602

1. Based on this description, what measure of center and spread should you report?

Since it is right skewed, we should report Median and IQR and Range

1. Find the mean, standard deviation, 5# summary, and IQR

Mean = 6.42 Min = 0.39

Std. Dev = 8.12 Q1= 0.398

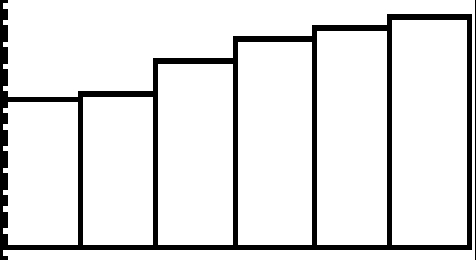
Med = 1.75

Q3 = 13

Max = 28

IQR = 12.602

1. Create a cumulative frequency histogram.



Cumulative

#

31

25

21

0 5 10 15 20 25 30

**SALARIES**

1. Heights (in cm) of 58 randomly selected Canadian students who participated in a survey

166.5 170 178 163 150.5 169 171 166 190 183 178 161

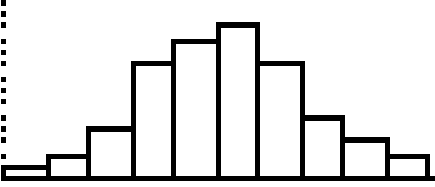
171 170 191 168.5 178.5 173 175 160.5 166 164 163 174

173 169 160 174 182 167 166 170 170 181 171.5 160

178 157 165 187 168 157.5 145.5 156 182 168.5 177 162.5

* 1. 185.5 151 159 177 171 176 177 181 186

1. Create a relative frequency histogram of the data. Describe the distribution.



**%**

20.7

15.5

6.9

1.7

145 150 155 160

**HEIGHTS**

Shape: roughly symmetric, unimodal

Center: Mean of 170.21

Spread: std. deviation of 9.9 and range of (145.5, 191)

1. Based on this description, what measure of center and spread should you report?

Since the distribution is roughly symmetric, we should report the mean and standard deviation.

1. Find the mean, standard deviation, 5# summary, and IQR

Mean = 170.21 Min = 145.5 Q3 = 177

Std. Dev = 9.9 Q1 = 163 Max = 191

IQR = 14 Med = 170

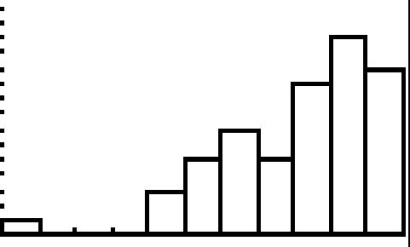
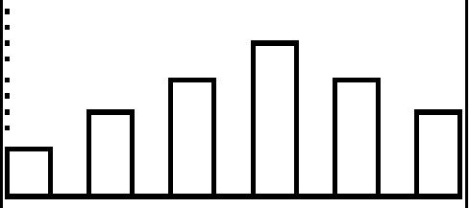
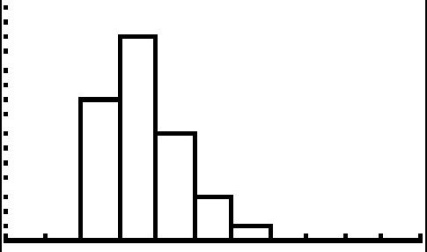
1. Use the following data. {30, 30, 30, 30, 30, 30, 30, 30}. Find the mean and standard deviation. Why is the standard deviation this value?

Mean = 30

Std. Deviatin = 0

The standard deviation is 0 because all the values are the same. The data does not “deviate” from the mean at all. So the average deviation = 0.

1. Describe the following distributions using the terms we learned in class. Scale on x-axis: (1, 12), bins = 1

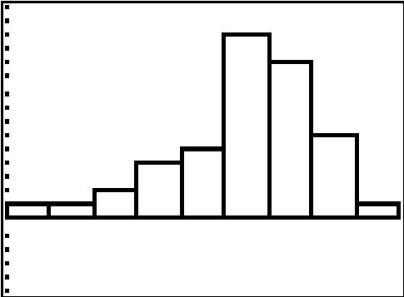
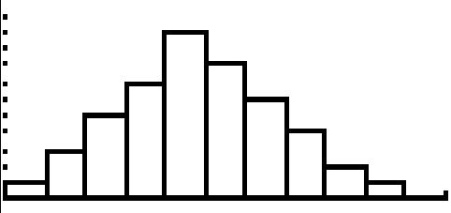
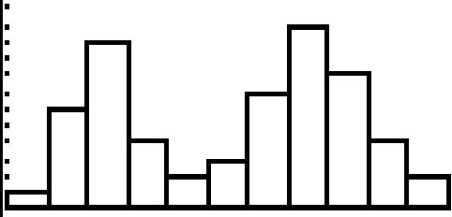
  

Shape: unimodal, left skew Shape: unimodal, symmetric shape: unimodal, right skewed

Center: approx. 8 center: approx. 7 center: approx. 4

Spread: (5, 11) spread: (1, 11) spread: (3, 7)

Outlier @ 1 granularity clustered

Shape: left skewed, unimodal shape: symmetric, unimodal shape: bimodal, symmetric

Center: approx. 5 center: approx. 5 center: approx. 6

Spread: (1, 10) spread: (1, 10) spread: (1, 11)

1. Use the following data: {20, 23, 24, 27, 29, 31, 30, 33, 36, 37, 35, 40}
2. Calculate the following statistics:

Mean 30.42

Median 30.5

Range (20, 40) = 20 units

IQR 10

Std. Dev. 6.127

1. Suppose we now add a new point to the data set: 60. Indicate whether adding the new point to the rest of the data made each of the summary statistics in part (a) increase, decrease, or stay about the same

Increase = mean, std. deviation, range

Same = median, IQR

1. A random sample of the heights of 24-34 year old women was taken (in inches). The following summary statistics were calculated.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistic** | **mean** | **st. dev.** | **min** | **Q1** | **med** | **Q3** | **max** |
| **Heights of 24-34 year old women** | **69.5** | **2.65** | **58** | **62** | **64** | **68** | **78** |

1. Based on the summary statistics would you describe the distribution as symmetric or skewed? Explain.

I would say the data are skewed because the mean is significantly greater than the median

1. Are there any outliers present? Show all work.

IQR = 68 – 62 = 6

1.5 x IQR = 9

UF = Q3 + 9 = 77

LF = Q1 – 9 = 53

Anything outside (53, 77) is considered an outlier, so 78 is an outlier.

**OR**

Mean + 2s = 69.5 + (2\*2.65) = (64.2, 74.8) Anything outside this range is an outlier, so 78 is an outlier.